# REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188



| 1. AGENCY USE ONLY ( Leave Blank) | 2. REPORT DATE<br>November 9, 2005 | 3. REPORT TYPE AND DATES COVERED<br>Final Progress Report, 15 Jul 02 - 14 July 05 |
|---|---|---|

4. TITLE AND SUBTITLE

Coordinated Anomaly Detection and Characterization in Wide Area Network Flows

5. FUNDING NUMBERS

DAAD19-02-1-0304

6. AUTHOR(S)

Professor Paul Barford

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Board of Regents – University of Wisconsin System
750 University Avenue
Madison, WI 53706-1490

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U. S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

43840.1-C1

11. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12 a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The ability to quickly and accurately identify anomalous behavior in computer networks is essential to assure that they perform efficiently, safely and reliably. The current standard in anomaly detection technology is autonomous packet level analysis that uses simple thresholds or rules to generate alerts. While these systems are effective in detecting and identifying some types of anomalous behavior, networks are still far from being robust or reliable. In this project, we are pursuing research initiatives aimed at developing the next generation of anomaly detection infrastructures, methods and tools.

Our initial efforts have focused in two areas - measurement and characterization of general types of anomalous traffic (misconfigurations, failures, flash crowds, etc.), and measurement and characterization of malicious network traffic (intrusions and attacks). Our focus in the former has been on applying multi-resolution analysis to IP flow data collected at our campus border router. Our focus in the latter has been on using intrusion data collected from a large number of networks to identify malicious activity. Both efforts have resulted in tools and systems that we will continue to develop.

Our future efforts will emphasize expansion and refinement of coordinated detection methods and wide deployment of these capabilities across the IPv4 address space as well as in the wireless domain.

14. SUBJECT TERMS

Network anomaly detection, network intrusion detection, empirical analysis of malicious network traffic, coordinated intrusion detection, automatic signature generation

15. NUMBER OF PAGES

8

16. PRICE CODE

| 17. SECURITY CLASSIFICATION OR REPORT | 18. SECURITY CLASSIFICATION ON THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| **UNCLASSIFIED** | **UNCLASSIFIED** | **UNCLASSIFIED** | **UL** |

NSN 7540-01-280-5500

**Standard Form 298 (Rev.2-89)**
Prescribed by ANSI Std. 239-18
298-102

# Final Progress Report

## 1. Forward

This final progress report for ARO grant #DAAD19-02-1-0304 entitled "Coordinated Anomaly Detection and Characterization in Wide Area Network Flows" was conducted over a three year period from July, 2002 to July, 2005. While the initial focus of the proposal was rather narrow, the study conducted over the period of the program was wide-ranging and multifaceted. Early on, the focus of the work broadened to become a comprehensive study of malicious activity in the Internet. Specific research projects were conducted on *(i)* characterizing the empirical behavior of malicious attacks and intrusions in the Internet, *(ii)* developing tools for measuring and evaluating attack and intrusion activity in the Internet, and *(iii)* developing tools to improve "Internet Situational Awareness". Nine technical papers were published in the most prestigious and highly selective Internet measurement and security related conferences as a direct result from support from this grant, and two US patents have been filed. Furthermore, the work has been widely cited and is formed the foundation for a large body of research activity both at the University of Wisconsin as well as other universities. Finally, the work that was supported from this grant has been presented in a wide variety of venues including research labs, universities, workshops and conferences.

## 2. Statement of Problem Studied

The ability to quickly and accurately identify anomalous behavior in computer networks is essential to assure that they perform efficiently, safely and reliably. The current standard in anomaly detection technology is autonomous packet level analysis that uses simple thresholds or rules to generate alerts. While these systems are effective in detecting and identifying some types of anomalous behavior, networks are still far from being robust or reliable. In this project, we originally set out to pursue research initiatives aimed at developing the next generation of anomaly detection infrastructures, methods and tools. Shortly after beginning this work, we expanded the scope of the problem space to include empirical measurement and analysis of malicious traffic in the Internet and the development of tools and systems to provide "Internet Situational Awareness", the ability to scale the perspective on malicious activity to a desired/required level. Prior to our work, the canonical systems used for this activity were network intrusion detection systems (NIDS) that have many known problems but served as a starting point for some of our efforts.

## 3. Summary of Most Important Results

This grant supported research and development activities that are summarized as follows:

1. *Wavelet-based Anomaly Detection.* We developed the first method for applying wavelets to the problem of statistical anomaly detection in network flow data. Wavelets are powerful tools for isolating discontinuities in both space and time. Using a unique labeled data set collected at the University of Wisconsin, we developed a method for applying wavelets to flow data collected from our campus border router. Our results show that our wavelet-based tool is extremely effective at isolating anomalies. We continue to develop these techniques in follow-on work.

2. *Global Characteristics of Internet Attacks and Intrusions.* Using the firewall and intrusion logs from over 1,700 networks worldwide (provided by Dshield.org), we conducted the first global analysis of Internet intrusion and attack activity. Our results show that these activities take place on a massive scale and that there is an increasing trend in the data. The results also laid the groundwork for our DOMINO project that is described below. A follow-on study was conducted using data collected from our iSink honeypot system (described below). This study coined the term "Internet Background Radiation" for the unwanted malicious traffic that courses through the Internet on a daily basis as a result of worms and other malware.

3. *Distributed Overlay for Monitoring InterNet Outbreaks* (DOMINO). DOMINO is a multifaceted system designed to use intrusion data from many collaborating sites to generate intrusion alerts in an accurate and timely fashion. The system is based on using peer-to-peer technology to facilitate participation. We have developed the architecture for DOMINO and have evaluated its effectiveness analytically from the perspective of false positive/negatives and on response time to identify new worm outbreaks. Our results show DOMINO to provide vastly superior capabilities when compared to intrusion monitoring systems deployed in isolation. Through support from the ARO DURIP program, we are in the process of building an instance of DOMINO in the live Internet.

4. *Internet Sink Monitors.* Internet Sink's (iSink) are packet monitors deployed on unused but routed IP address space that include the ability to respond to incoming connection requests. These so-called active honeypots provide a unique and extremely valuable perspective on malicious activity. The unique capability of iSink is that it is active response capability is scalable and is not based on virtual machines as are most other honeypots. We are able to monitor an entire class A network (16 million addresses) on a single PC. ISink data has been used in a number of our papers and continues to be a key source of data for our on-going activities. We also filed a US patent on iSink's technology.

5. *A Framework for Malicious Workload Generation.* Any system that is developed to detect attacks and intrusions needs to be thoroughly tested. Any such tests must be based on known ground truth in order to fully assess the overall effectiveness with respect to false alarms (both positive and negative). We developed an architectural framework for malicious workload generation that enables the flexible composition of malicious traffic such that both known attacks (such as the Welchia worm) and new attack variants can be realized. We realized this framework in a tool we call MACE

which we have enhanced with critical the ability to generate representative (from the perspective of both packet headers and payloads) benign traffic as well.

6. *Automatic Semantic-Aware Intrusion Signature Generation.* One of the most significant problems with current NIDS is that the signatures they used to detect malicious attacks all have to be crafted by hand after a new attack has been recognized. This process results in NIDS that have extremely high false alarm rates. We developed a process for automatically generating IDS signatures using data collected from iSink systems. This process is realized in a tool we call Nemean that is currently a prototype that can be used for off line tests. In our evaluation of Nemean's semantic-aware signatures showed that they have extremely high detection rates with almost no false positive alerts and do significantly better than the canonical Snort tool in similar tests. We have filed a US patent on Nemean technology and intend to work toward deploying this capability in ARL.

## Listing of Publications

### Papers published in peer-reviewed journals

None

### Papers published in peer-reviewed conference proceedings


1. Barford, Paul; Kline, Jeffery; Plonka, David; Ron, Amos. "A Signal Analysis of Network Traffic Anomalies," In Proceedings of ACM SIGCOMM Internet Measurement Workshop, November, 2002.

2. Yegneswaran, Vinod; Barford, Paul; Ullrich, Johannes. "Internet Intrusions: Global Characteristics and Prevalence", In Proceedings of ACM SIGMETRICS, 2003.

3. Yegneswaran, Vinod; Barford, Paul; Jha, Somesh. "Global Intrusion Detection in the DOMINO Overlay System" In Proceedings of ISOC Network and Distributed Systems Security Symposium (NDSS '04), February, 2004.

4. Yegneswaran, Vinod; Barford, Paul; Plonka, David. "On the Design and Use of Internet Sinks for Network Abuse Monitoring", In Proceedings of Symposium on Recent Advances in Intrusion Detection (RAID), September, 2004.

5. Pang, Rouming; Yegneswaran, Vinod; Barford, Paul; Paxson, Vern; Peterson, Larry. *Characteristics of Internet Background Radiation*, In Proceedings of ACM Internet Measurement Conference, October, 2004.

6. Sommers, Joel; Yegneswaran, Vinod; Barford, Paul. *A Framework for Malicious Workload Generation*, In Proceedings of ACM Internet Measurement Conference, October, 2004.

7. Barford, Paul; Jha, Somesh; Yegneswaran, Vinod. *Fusion and Filtering in Distributed Intrusion Detection Systems*, In Proceedings of the 42nd Annual Allerton Conference on Communication, Control and Computing, September, 2004.

8. Yegneswaran, Vinod; Giffin, Jon; Barford, Paul; Jha, Somesh. *An Architecture for Generating Semantic-aware Signatures*, In proceedings of USENIX Security Symposium, August, 2005.

9. Yegneswaran, Vinod; Sommers, Barford, Paul; Paxson, Vern. *Using Honeynets for Internet Situational Awareness*, In Proceedings of the ACM/USENIX Fourth Workshop on Hot Topics in Networks (Hotnets IV), November, 2005.

**Papers presented at meetings but not published in conference proceedings**

None

**Manuscripts submitted but not published**

10. Sommers, Joel; Yegneswaran, Vinod; Barford, Paul. *Toward Comprehensive Traffic Generation for Online IDS Evaluation*, Submitted for publication to ACM SIGMETRICS, 2006.

**Technical reports submitted to ARO**

None

## Listing of all Scientific Personnel

1. Paul Barford (PI)

2. Somesh Jha (co-PI)

3. Vinod Yegneswaran (student who will complete his PhD. in May, 2006)

4. Adam Smith (student in good standing in the department)

## Report of Inventions

1. "A Semantic-Aware Network Intrusion Signature Generator", filed in December, 2004.

2. "Scalable Monitor for Malicious Network Traffic", filed in August, 2005.